Pak. J. Statist.
2019 Vol. 35(1), 11-24

OPTIMUM STRATIFICATION BY TWO STRATIFYING VARIABLES USING MATHEMATICAL PROGRAMMING

Faizan Danish[§] and S.E.H. Rizvi
Division of Statistics and Computer Science, Faculty of Basic Sciences
Sher-e-Kashmir University of Agricultural Sciences and Technology
J, Main Campus, Chatha, Jammu-180009 (J&K), India
[§]Corresponding author Email: danishstat@gmail.com

ABSTRACT

In this paper, we have proposed an approach for obtaining optimum strata boundaries for one dependent variable with two auxiliary variables. The problem is formulated as non-linear programming problem which is solved by using dynamic programming. While comparing the proposed approach to the approaches already developed for single study variable and single auxiliary variable more precision is obtained in case of the proposed approach. Empirical study has been given in which each of the auxiliary variable is supposed to follow different distribution.

Mathematical Classification: 62D05

KEYWORDS

Dynamic programming, Nonlinear programming problem, Multistage decision problem.

1. INTRODUCTION

To use the stratified random sampling, choosing the best boundary points that make the strata internally homogeneous and the variances of the estimator within the strata be as small as possible is one of most need for the sampler. However sometimes when a single characteristic is under study and its frequency distribution is known, one could use this information effectively to achieve the best strata boundaries. If the frequency distribution of the study variable is unknown, it may be approximated from the past experience or using some prior knowledge obtained at a recent study. This problem was first considered by Dalenius (1950). He presented a set of minimal equations that could be solved for finding optimum boundary points. These equations could not usually be solved unless the number of strata is small. Hence attempts have been made by many authors such as Dalenius and Gurney (1951), Mahalanobis (1952), Aoyama (1954), Ekman (1959), Dalenius and Hodges (1959), Sethi (1963), Hess, *et al.* (1966), Singh and Sukhatme (1969, 1973), Singh (1977), Yadava and Singh (1984), etc. For determining strata boundaries majority of the authors have suggested approaches and obtain the calculus equations in terms of mean and variance of the stratum, but these equations are not of practical use as both mean and variance of the stratum are dependent of the boundary points due to which different rules were suggested to obtain approximately optimum strata boundaries.

Buhler and Deutler (1975) formulated the problem of determining OSB as an optimization problem and developed a computational technique to solve the problem using dynamic programming. Khan, *et al*. (2002) considered the problem of finding OSB as an equivalent problem of determining Optimum Strata Width (OSW) using Mathematical Programming approach. Later Khan, *et al.* (2005) formulated the problem of OSB for exponential study variable as an MPP and determined the optimum boundary points using dynamic programming. Khan *et al*. (2015) determined optimum strata boundaries and sample sizes for skewed population with log normal distribution.

For varying cost of every unit Danish *et al*. (2017) discussed the way of obtaining optimum strata boundaries. Danish and Rizvi (2017) studied problem of finding OSB by taking into consideration as the problem of optimum strata width (OSW), using MPP by dynamic programming technique, when the study variable is uniformly distributed.

In this paper, we have proposed an approach for obtaining OSB having single study variable with two auxiliary variables used as stratification variables. Furthermore the efficiency of the proposed approach have been calculated while comparing with several exiting methods.

2. FORMULATION OF PROBLEM AS MPP

For estimating the mean of the study variable when the population consisting of N units be divided into $L \times M$ strata based on two stratification variables X and Z, such that each stratum is homogenous within itself and heterogeneous between strata with respect to the character under study. Let N_{hk} be the units in the $(h,k)^{th}$ stratum, such that $\sum_{h=1}^{L}\sum_{k=1}^{M} N_{hk} = N$.

Let us assume n_{hk} be the sample size drawn from $(h,k)^{th}$ stratum such that $\sum_h\sum_k n_{hk} = n$. The population total can be expressed as $y = \sum_h\sum_k\sum_i y_{hki}$, where y_{hki}, $(i = 1,2,3,...,N_{hk})$ denotes the population unit in the $(h,k)^{th}$ stratum. $\bar{y}_{st} = \sum_h\sum_k W_{hk}\bar{y}_{hk}$ is expressed as an unbiased estimator of the population mean $\bar{Y}_N$ under stratified random sampling, where , $\bar{y}_{hk} = \frac{1}{n_{hk}}\sum_i y_{hki}$ and 'W_{hk}' denotes the weight of the $(h,k)^{th}$ stratum and is equal to $\frac{N_{hk}}{N}$.

The sampling variance of the unbiased estimator $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \sum_h\sum_k\left(\frac{1}{n} - \frac{1}{N}\right)W_{hk}^2\sigma_{hky}^2$$

This can be written as if the finite population correction (f.p.c) is ignored:

$$V\left(\bar{y}_{st}\right)=\sum_h\sum_k\frac{W_{hk}^2\sigma_{hky}^2}{n} \tag{1}$$

Let the regression model of study variable on auxiliary variables is of the form as

$$Y=\psi\left(x,z\right)+e \tag{2}$$

where, $\psi\left(x,z\right)$ be a linear function of 'X' and 'Z' and 'e' denotes the error term such that conditional expectation and variance of 'e' given X and Z be zero and $\phi\left(x,z\right)$ respectively.

Under model (2) the stratum mean 'μ_{hky}' and the stratum variance 'σ_{hky}^2' can be written as (Singh and Sukhatme 1969)

$$\mu_{hky}=\mu_{hk\psi} \tag{3}$$

and

$$\sigma_{hky}^2=\sigma_{hk\psi}^2+\mu_{hk\phi} \tag{4}$$

where $\mu_{hk\psi}$ and $\mu_{hk\phi}$ are the expected value of $\psi\left(x,z\right)$ and $\phi\left(x,z\right)$ respectively and $\sigma_{hk\lambda}^2$ denotes the variance of $\lambda(x,z)$ in the $\left(h,k\right)^{th}$ stratum. σ_{hky}^2 can be expressed as

$$\sigma_{hky}^2=\sigma_{hk\lambda}^2+\sigma_{hk\varepsilon}^2 \tag{5}$$

(Dalenius and Gurney 1951) if ψ and e are not correlated.

Let $f\left(x,y,z\right)$, $f\left(x,z\right)$ and $f\left(x\right)$ & $f\left(z\right)$ be joint density function of $\left(X,Y,Z\right)$, joint marginal density function of X and Z and the frequency function of the auxiliary variables X and Z respectively defined in the interval [a, b] and [c, d].

If the population mean of the study variable 'Y' is estimated under the variance given in equation (1), then the problem of determining the strata boundaries is to cut up the ranges $d_x=b-a$ and $t_z=d-c$, at (L-1) and (M-1) intermediate points as

$$a=x_0\le x_1\le\ldots\le x_{L-1}\le x_L=b$$

and

$$c=z_0\le z_1\le\ldots\le z_{M-1}\le z_M=d$$

respectively such that the equation (2) is minimum.

For minimizing (1) it is sufficient to minimize $\sum_h\sum_k W_{hk}^2\sigma_{hky}^2$. Thus while using (4), we have

$$\sum_h\sum_k W_{hk}^2\left(\sigma_{hk\lambda}^2+\mu_{hk\phi}\right) \tag{6}$$

If the functions $f(x,z), \psi(x,z)$ and $\phi(x,z)$ are integrable then, we can write

$$\sigma_{hk\lambda}^2 = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda^2(x,z) f(x,z) \partial x \partial z - \mu_{hk\lambda}^2 \tag{7}$$

and

$$\mu_{hk\phi} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \phi(x,z) f(x,z) \partial x \partial z \tag{8}$$

where,

$$\mu_{hk\lambda} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} \lambda(x,z) f(x,z) \partial x \partial z \tag{9}$$

and

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f(x,z) \partial x \partial z \tag{10}$$

Let

$$\psi(x_h, x_{h-1}, z_k, z_{k-1}) = W_{hk}^2 \left(\sigma_{hk\lambda}^2 + \mu_{hk\phi}\right) \tag{11}$$

and

$$d_x = b - a = x_L - x_0 \tag{12}$$

$$t_z = d - c = z_M - z_0 \tag{13}$$

Then, in the bivariate stratification a problem of determining the strata boundaries (x_h, z_k) is to break up the ranges of (12) and (13) at intermediate points in order to estimate $x_1 \le x_{2\le} ... \le x_{L-2} \le x_{L-1}$ and $z_1 \le z_{2\le} ... \le z_{M-2} \le z_{M-1}$. Then, the reasonable criterion for determining optimum strata boundaries (OSB) (x_h, z_k) is to minimize

Minimize $\sum_h \sum_k \psi_{hk}(x_h, x_{h-1}, z_k, z_{k-1})$

Subject to

$$a = x_0 \le x_1 \le ... \le x_{L-1} \le x_L = b \tag{14}$$
$$c = z_0 \le z_1 \le ... \le z_{M-1} \le z_M = d$$

and

$$\sum_h \sum_k n_{hk} = n$$

Let $V_h = x_h - x_{h-1}$ and $U_k = z_k - z_{k-1}$ denotes the total length or width of the $(h,k)^{th}$ stratum. Then, using (12) and (13), the ranges can be expressed as

$$\sum_h V_h = \sum_h (x_h - x_{h-1}) = b - a = d_x \tag{15}$$

$$\sum_k U_k = \sum_k (z_k - z_{k-1}) = d - c = t_z \tag{16}$$

Using the concept of dynamic programming for two dimensional, the of two way optimum stratification can be expressed as

$$Min \quad \sum_h \sum_k \psi_{hk}\left(x_h, x_{h-1}, z_k, z_{k-1}\right)$$

Subject to

$$\begin{aligned} &\left(x_h, z_k\right) = \left(x_{h-1} + V_h, z_{k-1} + U_k\right) \\ &\left(x_h, z_k\right) \in [a,d] \times [c,d] \\ &\left(V_h, U_k\right) \in B_h\left(x_{h-1}\right) \times B_k\left(z_{k-1}\right) \\ &\qquad = \left[0, b - x_{h-1}\right] \times \left[0, d - z_{k-1}\right] \\ &h = 1,2,...,L \quad and \quad k = 1,2,...,M \end{aligned} \tag{17}$$

Let $\psi^*_{x_h}\left(x_{h-1}, z^{i-1}\right)$ be the optimal value for the objective function (14) for the strata (h,k) to (L,k) for all $k = 1,2,...,M$ given that the lower bound for the strata (h,k) for $k = 1,2,...,M$ is x_{h-1}. The functional equation of Bellman with respect to the first part of the i[th] iteration is then given by

$$\psi^*_{x_h}\left(x_{h-1}, z^{i-1}\right) = \underset{V_h \in B_h\left(x_{h-1}\right)}{Minimize} \left\{ \sum_{k=1}^{M} \psi\left(x_{h-1}, x_h, z^{i-1}_{k-1}, z^{i-1}_k\right) + \psi^*_{x_{h+1}}\left(x_h, z^{i-1}\right) \middle| x_h = x_{h-1} + V_h \right\}$$

The OSB for the first part of the i[th] iteration are given by $\left(x^i, z^{i-1}\right)$. For the second part of the i[th] iteration, the points of stratification x^i are in turn considered as fixed. Restating the problem of determining OSB as the problem of determining optimum points $\left(V_h, U_k\right)$, adding equation (15) and (16) as a constraint, the problem (14) can be treated as an equation problem of determining OSW, $V_1, V_2, ..., V_L$ and $U_1, U_2, ..., U_M$ and is expressed as the following MPP:

$$\text{Minimize} \quad \sum_h \sum_k \psi_{hk}\left(x_h, x_{h-1}, z_k, z_{k-1}\right)$$

Subject to

$$\begin{aligned} &\sum_h V_h = d_x \\ &\sum_k U_k = t_z \;, h = 1,2,...,L \quad and \quad k = 1,2,...,M \end{aligned} \tag{18}$$

and

$$V_h \geq 0 \text{ and } U_k \geq 0$$

For the two auxiliary variables x_0 and z_0 are known of *X* and *Z* respectively. The first term $\psi_{11}\left(x_1, x_0, z_1, z_0\right)$ in the objective function (18) is the function of $\left(V_1, U_1\right)$ alone provided the $\left(V_1, U_1\right)$ is known, the second term $\psi_{22}\left(x_2, x_1, z_2, z_1\right)$ will be the function of $\left(V_2, U_2\right)$ alone and so on. The special nature of the above function MPP (18) may be treated as the function of $\left(V_h, U_k\right)$ and can be expressed as

$$
\begin{aligned}
&\text{Minimize } \sum_h \sum_k \psi_{hk}\left(V_h, U_k\right) \\
&\text{Subject to} \\
&\qquad \sum_h V_h = d_x \\
&\qquad \sum_k U_k = t_z \,,\; h = 1,2,\ldots,L \;\; and \;\; k = 1,2,\ldots,M \\
&\text{and} \\
&\qquad V_h \geq 0 \text{ and } U_k \geq 0 .
\end{aligned}
\tag{19}
$$

3. THE SOLUTION PROCEDURE

The MPP defined in (19) is having objective function and constraints as separable functions of $\left(V_h, U_k\right)$, then it is the case of multistage decision problem that allows us to use dynamic programming technique. A dynamic programming model is generally a recursive equation and the recursive equation links to different stages of the problem.

Consider $\left(L_1 \times M_1\right)$ be the first sub problem of equation (19) i.e. $L_1 < L, M_1 < M$

$$
\begin{aligned}
&\text{Minimize } \sum_{h=1}^{L_1} \sum_{k=1}^{M_1} \psi_{hk}\left(x_{h-1}, x_h, z_{k-1}, z_k\right) \\
&\text{Subject to} \\
&\qquad \sum_{h=1}^{L-1} V_h = d_{L_1} \\
&\qquad \sum_{k=1}^{M-1} U_k = t_{M_1} \,,\; h = 1,2,\ldots,L_1 \;\; and \;\; k = 1,2,\ldots,M_1 \\
&\text{and} \\
&\qquad V_h \geq 0 \text{ and } U_k \geq 0 \\
&\text{where } \; d_{L_1} < V, t_{M_1} < M
\end{aligned}
\tag{20}
$$

The transformation functions are given by

$$
\begin{aligned}
d_{L_1} &= V_1 + V_2 + \ldots + V_{L_1} \\
d_{L_1-1} &= V_1 + V_2 + \ldots + V_{L_1-1} = d_{L_1} - V_{L_1} \\
d_{L_1-2} &= V_1 + V_2 + \ldots + V_{L_1-2} = d_{L_1-1} - V_{L_1-1} \\
&\;\;\vdots \\
d_2 &= V_1 + V_2 = d_3 - V_3 \\
d_1 &= V_1 = d_2 - V_2
\end{aligned}
$$

Similarly, we have

$$t_{M_1} = U_1 + U_2 + \ldots + U_{M_1}$$
$$t_{M_1-1} = U_1 + U_2 + \ldots + U_{M_1-1} = t_{M_1} - U_{M_1}$$
$$t_{M_1-2} = U_1 + U_2 + \ldots + U_{M_1-2} = t_{M_1-1} - U_{M_1-1}$$
$$\vdots$$
$$t_2 = U_1 + U_2 = t_3 - U_3$$
$$t_1 = U_1 = t_2 - U_2$$

Let

$$\psi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = Min\left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1}\psi_{hk}\left(V_h, U_k\right)\middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1-1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1-1}\right]$$
$$and \quad V_h \geq 0, U_k \geq 0; h = 1,2,3,\ldots,L_1 \quad and \quad k = 1,2,3,\ldots,M_1$$

denotes the minimum value of objective function of equation (20).With this definition of $\psi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$, the MPP (19) is equivalent to finding $\psi_{L \times M}\left(d_x, t_z\right)$ recursively by defining $\psi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right)$ for $L_1 = 1,2,\ldots,L$ *and* $M_1 = 1,2,\ldots,M$; $0 \leq d_{L_1} \leq V$, $0 \leq t_{M_1} \leq U$.

$$\psi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = Min\left[\begin{array}{l}\psi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right) + \left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1}\psi_{hk}\left(V_h, U_k\right)\right| \\ \qquad \left.\sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1}\right]\end{array}\right] \quad (21)$$
$$and \quad V_h \geq 0, U_k \geq 0; h = 1,2,3,\ldots,L_1 \quad and \quad k = 1,2,3,\ldots,M_1$$

For fixed value of $\left(V_{L_1}, U_{M_1}\right)$, $0 \leq d_{L_1} \leq V$, $0 \leq t_{M_1} \leq U$.

$$\psi_{L_1 \times M_1}\left(d_{L_1}, t_{M_1}\right) = \psi_{L_1 \times M_1}\left(V_{L_1}, U_{M_1}\right) + Min\left[\sum_{h=1}^{L_1-1}\sum_{k=1}^{M_1-1}\psi_{hk}\left(V_h, U_k\right)\middle| \sum_{h=1}^{L_1-1} V_h = d_{L_1} - V_{L_1}, \sum_{k=1}^{M_1-1} U_k = t_{M_1} - U_{M_1}\right]$$

and

$$V_h \geq 0, \quad h = 1,2,\ldots,L_1 - 1$$
$$U_k \geq 0, \quad k = 1,2,\ldots,M_1 - 1$$

and

$$1 \leq L_1 \leq L, \quad 1 \leq M_1 \leq M$$

Following the above procedure, we can have the recursive equation of dynamic programming for obtaining OSB. Let us assume the linear regression model be:

$$Y = \alpha + \beta x + \gamma z + \varepsilon$$

then for the case of independence between error term and the two auxiliary variables, we have

$$\sigma_{hky}^2 = \beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2 \tag{22}$$

$$\sigma_{hkx}^2 = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x^2 f(x) \partial x \partial z - \mu_{hkx}^2 \tag{23}$$

$$\sigma_{hkz}^2 = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z^2 f(z) \partial z \partial x - \mu_{hkz}^2 \tag{24}$$

where

$$W_{hk} = \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} f(x,z) \partial x \partial z \tag{25}$$

$$\mu_{hkx} = \frac{1}{W_{hk}} \int_{z_{k-1}}^{z_k} \int_{x_{h-1}}^{x_h} x f(x) \partial x \partial z \,,\ \mu_{hkz} = \frac{1}{W_{hk}} \int_{x_{h-1}}^{x_h} \int_{z_{k-1}}^{z_k} z f(z) \partial z \partial x$$

4. EMPIRICAL STUDY

Let us consider that one of the auxiliary variables ‘X’ follows a distribution with pdf as

$$f(x) = \begin{cases} e^{-x+1} & ; \quad 1 \le z \le 4 \\ 0 & ; otherwise \end{cases} \tag{26}$$

and the other variable ‘Z’ follows Pareto distribution with pdf as

$$f(z) \begin{cases} \dfrac{ab^a}{z^{a+1}} & ; \quad z \in (b, \infty) \\ 0 & ; \quad otherwise \end{cases} \tag{27}$$

where $a > 0$ is the shape of the parameter and $b > 0$ is the scale parameter.

In order to obtain OSB, when the auxiliary variables X and Z have distribution functions given in (26) and (27), we need to find the values of (23)-(25), we have

$$W_{hk} = b^a e^{-x_{h-1}+1} \left(1 - e^{-V_h}\right) \left[\frac{\left(U_k + z_{k-1}\right)^2 - \left(z_{k-1}\right)^a}{\left(U_k + z_{k-1}\right)^2 \left(z_{k-1}\right)^a} \right] \tag{28}$$

$$\sigma_{hkx}^2 = U_k V_h \left(U_k + z_{k-1}\right)^a \left(z_{k-1}\right)^a \left\{ \frac{\frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[\left(V_h + x_{h-1}\right)\left(V_h\left(V_h + x_{h-1}\right) + x_{h-1}\left(1 + x_{h-1}\right)\right)\right]}{b^a e^{-x_{h-1}+1}\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^a - \left(z_{k-1}\right)^a\right]} \right\}$$
$$- \left(\frac{U_k\left(U_k + z_{k-1}\right)^2 \left(z_{k-1}\right)^a \left[x_{h-1} - \left(V_h + x_{h-1} + 1\right)e^{-V_h} + 1\right]}{b^a\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^a - \left(z_{k-1}\right)^a\right]} \right)^2 \quad (29)$$

$$\sigma_{hkz}^2 = \frac{a\left(U_k + z_{k-1}\right)^2 \left(z_{k-1}\right)^2 \left[\left(U_k + z_{k-1}\right)^{a-2} - \left(z_{k-1}\right)^{a-2}\right]}{\left(a-2\right)e^{-x_{h-1}+1}\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^a - \left(z_{k-1}\right)^a\right]}$$
$$- \left(\frac{a\left(U_k + z_{k-1}\right)\left(z_{k-1}\right)\left[\left(U_k + z_{k-1}\right)^{a-1} - \left(z_{k-1}\right)^{a-1}\right]}{\left(a-1\right)e^{-x_{h-1}+1}\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^a - \left(z_{k-1}\right)^a\right]} \right)^2 \quad (30)$$

Using the variance formula given in (1) and substitute obtained in (22), we get MPP as

$$\text{Minimize } \sum_h \sum_k W_{hk}^2 \left(\beta^2 \sigma_{hkx}^2 + \gamma^2 \sigma_{hkz}^2\right)$$

Subject to

$$\sum_h V_h = d_x$$
$$\sum_k U_k = t_z \quad (31)$$
$$\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{matrix} h = 1, 2, \ldots, L \\ k = 1, 2, \ldots, M \end{matrix}$$

Substituting (28)-(30), we have

Minimize

$$\sum_h\sum_k\left[b^a e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\frac{\left(U_k+z_{k-1}\right)^2-\left(z_{k-1}\right)^a}{\left(U_k+z_{k-1}\right)^2\left(z_{k-1}\right)^a}\right]\right]^2$$

$$\left\{\begin{array}{l}\beta^2 U_k V_h\left(U_k+z_{k-1}\right)^a\left(z_{k-1}\right)^a\\ \left\{\dfrac{\frac{4}{3}\left(V_h^2+3x_{h-1}^2+3V_h x_{h-1}\right)-\frac{1}{2}\left[\left(V_h+x_{h-1}\right)\left(V_h\left(V_h+x_{h-1}\right)+x_{h-1}\left(1+x_{h-1}\right)\right)\right]}{b^a e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k+z_{k-1}\right)^a-\left(z_{k-1}\right)^a\right]}\right\}\end{array}\right.$$

$$-\left(A_1\right)^2+\gamma^2\frac{a\left(U_k+z_{k-1}\right)^2\left(z_{k-1}\right)^2\left[\left(U_k+z_{k-1}\right)^{a-2}-\left(z_{k-1}\right)^{a-2}\right]}{\left(a-2\right)e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k+z_{k-1}\right)^a-\left(z_{k-1}\right)^a\right]}$$

$$\left.-\left(\frac{a\left(U_k+z_{k-1}\right)\left(z_{k-1}\right)\left[\left(U_k+z_{k-1}\right)^{a-1}-\left(z_{k-1}\right)^{a-1}\right]}{\left(a-1\right)e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\left(U_k+z_{k-1}\right)^a-\left(z_{k-1}\right)^a\right]}\right)^2\right\}$$

Subject to

$$\begin{aligned}&\sum_h V_h=d_x\\&\sum_k U_k=t_z\\&\forall V_h\ge 0, U_k\ge 0 \quad , \quad \begin{array}{l}h=1,2,\dots,L\\k=1,2,\dots,M\end{array}\end{aligned}\qquad (32)$$

where

$$A_1=\frac{U_k\left(U_k+z_{k-1}\right)^2\left(z_{k-1}\right)^a\left[x_{h-1}-\left(V_h+x_{h-1}+1\right)e^{-V_h}+1\right]}{b^a\left(1-e^{-V_h}\right)\left[\left(U_k+z_{k-1}\right)^a-\left(z_{k-1}\right)^a\right]}$$

Assuming the variable Z that follows Pareto distribution is defined in the interval [1.000, 10.000] and also assume that a = 1.342 and this implies that b = 1.000421. The (32) MPP can be put as

Minimize

$$\sum_h\sum_k\left[(1.00056)e^{-x_{h-1}+1}\left(1-e^{-V_h}\right)\left[\frac{\left(U_k+z_{k-1}\right)^{1.342}-\left(z_{k-1}\right)^{1.342}}{\left(U_k+z_{k-1}\right)^{1.342}\left(z_{k-1}\right)^{1.342}}\right]\right]^2$$

$$\left\{\begin{array}{l}\beta^2 U_k V_h \left(U_k + z_{k-1}\right)^{1.342}\left(z_{k-1}\right)^{1.342} \\ \left\{\dfrac{\frac{4}{3}\left(V_h^2 + 3x_{h-1}^2 + 3V_h x_{h-1}\right) - \frac{1}{2}\left[\left(V_h + x_{h-1}\right)\left(V_h\left(V_h + x_{h-1}\right) + x_{h-1}\left(1 + x_{h-1}\right)\right)\right]}{(1.00056)e^{-x_{h-1}+1}\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]}\right.\\ \left. - \left(\dfrac{U_k\left(U_k + z_{k-1}\right)^{1.342}\left(z_{k-1}\right)^{1.342}\left[x_{h-1} - \left(V_h + x_{h-1} + 1\right)e^{-V_h} + 1\right]}{(1.00056)\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]}\right)^2\right\} \\ + \gamma^2 \dfrac{1.342\left(U_k + z_{k-1}\right)^2\left(z_{k-1}\right)^2\left[\left(U_k + z_{k-1}\right)^{-0.658} - \left(z_{k-1}\right)^{-0.658}\right]}{\left(-0.658\right)e^{-x_{h-1}+1}\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]} \\ - \left(\dfrac{1.342\left(U_k + z_{k-1}\right)\left(z_{k-1}\right)\left[\left(U_k + z_{k-1}\right)^{0.342} - \left(z_{k-1}\right)^{0.342}\right]}{\left(0.342\right)e^{-x_{h-1}+1}\left(1 - e^{-V_h}\right)\left[\left(U_k + z_{k-1}\right)^{1.342} - \left(z_{k-1}\right)^{1.342}\right]}\right)^2\end{array}\right\}$$

Subject to

$$\begin{aligned} &\sum_h V_h = 3 \\ &\sum_k U_k = 9 \qquad\qquad (33) \\ &\forall V_h \geq 0, U_k \geq 0 \quad , \quad \begin{array}{l} h = 1, 2, \ldots, L \\ k = 1, 2, \ldots, M \end{array} \end{aligned}$$

By using the pdf's given in (26) and (27) the simulation has been made to get the values of $\beta = 0.0743, \lambda = 0.421$ in R-Software. Solving the (33) following the recursive equations by executing a computer programme in LINGO to obtain OSB for the above objective function having total 12 (3×4) strata are presented in table below:

Table I
Stratification Points when the Auxiliary Variables X and Z follow Exponential and Pareto Distributions

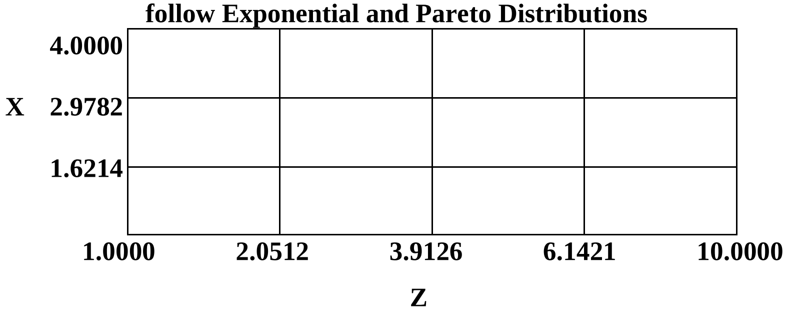

Table II
OSB and Variance for Exponential and Pareto Distributed Auxiliary Variables

OSB (x_h, z_k)	Variance
(1.6214,2.0512) (2.9782,2.0512) (4.0000,2.0512) (1.6214,3.9126) (2.9782,3.9126) (4.0000,3.9126) (1.6214,6.1421) (2.9782,6.1421) (4.0000,6.1421) (1.6214,10.0000) (2.9782,10.0000) (4.0000,10.0000)	0.002412

This shows the OSB when we have the total of 12 strata, 3 along *X* variable and 4 along *Z* variable as also the variance obtained by the proposed method. To illustrate it graphically the computation details of the proposed design, we consider the population of size 5000.

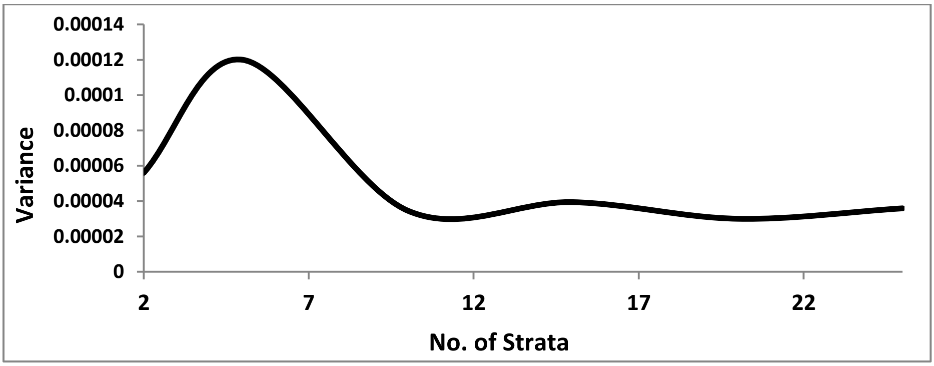

Fig. 1: Curve of Number of Starata and Corresponding Variance

From this graph, there is no substantial gain in efficiency for more than 20 strata. Thus the sufficient strata boundaries for each of the auxiliary variables would be 4 or 5.

Now in order to make the comparison of the proposed methods and the existing methods a simulation study is carried out to discuss the efficiency of the proposed method .The existing methods that are to be considered for comparison are:

1) Dalenius and Gurney (1951) method
2) Gunning and Horgan (2004) geometric method
3) Khan *et al*. (2015)
4) Proposed method

For making comparisons of these methods let us generate a population of size N = 10000 by using the two distributions using the R-software. By assuming 3 strata along the X axis and 3 along the z-axis, then the OSB's are determined by using R-package 1-2 and LINGO for 3 and 4. We get

Table III
Variances obtained through Different Methods

Method of Stratification	$v(\bar{x}_{st})$ (in e -07)	$v(\bar{z}_{st})$ (in e -07)	Total variance (in e-07)
Dalenius and Gurney (1951) method	498.7265	159.0518	657.7783
Gunning and Horgan (2004) geometric method	2089.6193	1568.9275	2474.6664
Khan *et al*. (2015)	394.6148	385.0471	779.6619
Proposed method	307.6284	127.0637	434.6921

Thus, it reveals that the proposed method of stratification shows maximum reduction in variance of both the estimates as compared to other method. Thus, the proposed bi-variate stratification method is a better option for obtaining the OSB

5. CONCLUSION

In this paper we have proposed an approach for obtaining optimum strata boundaries when a single study is of our interest for determining the two dimensional Optimum Strata Boundaries (OSB) when more than one having two stratification variables closely related to the study variable. However while solving the problem is formulated in a mathematical programming form that too is then solved using the concept of dynamic programming approach. Furthermore, the proposed method is compared with Dalenius and Gurney (1951) method, Gunning and Horgan (2004) geometric method and Khan *et al*. (2015) which results in more efficiency than all these methods. Thus we can conclude that using single study variable with two auxiliary is more preferred than using single study variable with single auxiliary variable.

6. ACKNOWLEDGEMENTS

The authors are grateful to the learned referee for their valuable comments and suggestions that helped in improving this paper. Further, we are highly thankful to the Editor of Pakistan Journal of Statistics and its Technical Editor Dr. Zahoor for their kind efforts.

7. REFERENCES

1. Aoyama, H. (1954). A study of stratified random sampling. *Ann. Inst. Stat. Math.*, 6, 1-36.
2. Bellman, R.E. (1957). *Dynamic Programming*. Princetown University Press, New Jersey.

3. Bhler, W. and Deutler, T. (1975). Optimal stratification and grouping by dynamic programming. *Metrika*, 22, 161-175.
4. Dalenius, T. (1950). The problem of optimum stratification-II. *Skand. Aktuartidskr*, 33, 203-213.
5. Dalenius, T. and Gurney, M. (1951). The problem of optimum stratification. *Skand. Akt.*, 34, 133-148.
6. Dalenius, T. and Hodges, J.L. (1959). Minimum variance stratification. *J. Amer. Statist. Assoc.*, 54, 88-101.
7. Danish, F., Rizvi, S.E.H., Jeelani, M.I. and Reashi, J.A. (2017). Obtaining strata boundaries under proportional allocation with varying cost of every unit. *Pakistan Journal of Statistics and Operation Research*, 13(3), 567-574.
8. Danish, F., Rizvi, S.E.H., Sharma, M.K. and Jeelani, M.I. (2017). Optimum Stratification Using Mathematical Programming Approach: A Review. Journal of Statistics Applications & Probability Letters, 4(3), 123-129.
9. Ekman, G. (1959). Approximate expression for conditional mean and variance over small intervals of a continuous distribution. *Ann. Inst. Stat. Math.*, 30, 1131-1134.
10. Gunning, P. and Horgan, J.M. (2004). A new algorithm for the construction of stratum boundaries in skewed populations. *Survey Methodology*, 30(2), 159-166.
11. Hess, I., Sethi, V.K. and Balakrishnan, T.R. (1966). Stratification: A practical Investigation. *J. Amer. Statist. Assoc.*, 61, 71-90.
12. Khan, E.A., Khan, M.G.M. and Ahsan, M.J. (2002). Optimum stratification: A mathematical programming approach. *Culcutta Statistical Association Bulletin*, 52(special), 205-208.
13. Khan, M.G., Sehar, N. and Ahsan, M.J. (2005). Optimum stratification for exponential study variable under Neyman allocation. *Journal of the Indian Society of Agricultural Statistics*, 59(2), 146-150.
14. Khan, M.G.M., Dinesh, R., Ansari, A.H. and Ahsan, M.J. (2015). Determining optimum strata boundaries and sample sizes for skewed population with log-normal distribution. *Communication in Statistics-Simulation and Computation*. 44, 1364-1387.
15. Mahalanobis, P.C. (1952). Some Aspects of the Design of Sample Surveys. *Sankhya*, 12, 1-7.
16. Sethi, V.K. (1963). A note on optimum stratification of population for estimating the population mean. *Aust. J. Statist.*, 5, 20-23.
17. Singh, R. (1977). A note on optimum stratification for equal allocation with ratio and regression methods of estimation. *Aust. J. Statist.*, 19(2), 96-104.
18. Singh, R. and Sukhatme, B.V. (1969). Optimum stratification. *Ann. Inst. Stat. Math.*, 21, 515-528.
19. Singh, R. and Sukhatme, B.V. (1973). Optimum stratification with ratio and regression methods of estimation. *Ann. Inst. Stat. Math.*, 25, 627-633.
20. Yadava, S.S. and Singh, R. (1984). Optimum stratification for allocation proportional to strata totals for simple random sampling scheme. *Commun. Statist. Theor. Math.*, 13(22), 2793-2806.